UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/064,435	07/12/2002	Prateek Sarkar	111747	2489

27074	7590	04/04/2006

OLIFF & BERRIDGE, PLC.
P.O. BOX 19928
ALEXANDRIA, VA 22320

EXAMINER
SCHAFFER, JONATHAN C

ART UNIT	PAPER NUMBER
2624	

DATE MAILED: 04/04/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 10/03)

	Application No.	Applicant(s)
Office Action Summary	10/064,435	SARKAR ET AL.
	Examiner	Art Unit
	Jonathan C. Schaffer	2624

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on <u>18 January 2006</u>.

2a)☒ This action is **FINAL.**   2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-31* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-31* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All   b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
    Paper No(s)/Mail Date _____.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: _____.

DETAILED ACTION

Response to Arguments

1.   Applicant's response to the last Office Action, filed 10/18/2005, has been entered and made of record.

2.   Applicant has amended claims 1, 2, 10, 12, 17, 19 and 25-31. Claims 25-31 were previously rejected under 35 U.S.C. § 101 for reciting non-statutory subject matter and in view of the current amendments has overcome that rejection. Claims 1-31 are currently pending.

3.   Applicant's arguments with respect to claims 1-31 have been considered but are moot in view of the new ground(s) of rejection.

Claim Rejections - 35 USC § 102

4.   The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

5.   Claims 1-31 are rejected under 35 U.S.C. 102(e) as being anticipated by Jamali (U.S. Patent Number 6,269,188).

1. A method for automatic triage of a text passage outputted by an optical character recognition system, the OCR-output text passage having multiple OCR-output characters, the method comprising:

determining at least one OCR-output character attribute for each of the OCR-output characters in the OCR-output text passage;

Jamali discloses determining at least one OCR-output character attribute for each of the OCR-output characters in the OCR-output text passage in one embodiment shape is the attribute (col. 4, I. 25-27).

determining an error rate for the OCR-output text passage using a triage model and the determined OCR-output character attributes;

Fig. 2a & 5a

and comparing the determined error rate for the OCR-output text passage with an OCR-output text passage threshold error rate to perform an OCR-output text passage triage decision.

Fig. 5a

2. The method of claim 1, wherein determining an error rate for the OCR-output text passage comprises:

providing the OCR-output character attributes to the triage model;

The shape attribute is provided to the triage model, which is then transformed into an accuracy value (col. 4, I. 25-27 & Fig. 2a-c & Fig. 5a).

determining a character interpretation error value for each OCR-output character based on a probability of the at least one OCR-output character attribute being erroneously interpreted by the system;

Fig. 2b

and determining a text passage error value based on the at least one character interpretation error value determined for each OCR-output character.

Fig. 5a

3. The method of claim 2, further comprising:

determining a number representing a sum of OCR-output characters in the OCR-output text passage;

Fig. 5a (528)

and dividing the text passage error value by the number representing the sum of OCR-output characters.

Fig. 5a (532)

4. The method of claim 1, wherein determining at least one OCR-output character attribute for each OCR-output character comprises selecting the at least one OCR-output character attribute from a plurality of OCR-output character attributes.

Jamali discloses multiple attributes two of which are shape (col. 4, l. 25-27) and font (col. 5, l. 23-35).

5.   The method of claim 4, wherein the plurality of OCR-output character attributes includes at least one of a character class, a confidence descriptor class, a language of the text passage, a text passage publication date, a typeface in which the text passage is printed, an image-based feature of an individual character image and metadata attached to the text passage.

Jamali discloses multiple attributes two of which are shape (col. 4, l. 25-27) and font (col. 5, l. 23-35).

6.   The method of claim 1, wherein the text passage to be triaged includes at least one of pages, characters, words, phrases, text-lines, sentences, paragraphs, columns of text, blocks of text, text articles, multi-page documents, collections of single-page documents and collections of multi-page documents.

Jamali discloses word groupings (col. 4, l. 30).

7.   The method of claim 1, wherein the OCR-output text passage triage decision includes at least one of sending the OCR-output text passage directly to an end user without post-OCR processing, sending the OCR-output text passage through a post-OCR inspection and processing stage, and sending the original text passage image to be keyed in manually.

Fig. 2a

8.   The method of claim 1, wherein the triage model is a trained off-line triage model.

Jamali discloses stored data files such as binary, gray-scale, and color image files, a table 170, text files, and programs including a word accuracy calculation program and one or more OCR programs (col. 3, l. 56-59). Jamali's character accuracy value is determined by the differences in shape between the OCR-output character and the stored template character (col. 4, l. 25-27), which indicates that a trained off-line triage model is an intrinsic part of the system.

9. The method of claim 1, wherein the OCR-output text passage threshold error rate is a predetermined value.

Jamali discloses a system with a user-predefined threshold (col. 6, l. 8-11).

10. The method of claim 7, wherein sending the OCR-output text passage through the post-OCR inspection and processing stage comprises:

determining at least one text passage error probability value for each OCR-output text passage as a correction operator detects and corrects an error in the OCR-output text passage;

Fig. 7 & 8a

and alerting the correction operator when the at least one text passage error probability value is improved so as to meet the OCR-output text passage threshold error value, wherein the text passage error probability value for each OCR-output text passage is based on a probability of the respective OCR-output character attributes being erroneously interpreted by the system.

The word accuracy value is ultimately displayed to the operator (Fig. 2a) after all calculations are done and most accurate word has been determined to be greater than the threshold defined by the operator (Figs. 5d, 7, 8a & col. 6, l. 8-11).

11. The method of claim 10, wherein determining the text passage error probability value for an OCR-output text passage comprises:

determining OCR-output text passage error probability values for a plurality of selected portions of the OCR-output text passage;

Fig. 8a

and arranging the plurality of selected portions of the OCR-output text passage based on the determined OCR-output text passage error probability values such that the selected portions having the highest OCR-output text passage error probability values are displayed first to the correction operator.

Fig. 7 (736) & 9 (912, 916)

12. A computer-implemented method for triage of a plurality of OCR-output text passages, each OCR-output text passage having multiple OCR-output characters, the method comprising:

selecting a set of OCR-output character attributes from a plurality of OCR-output character attributes for each OCR-output character;

See the rejection of Claim 1, first limitation.

determining an OCR-output character error value for each OCR-output character based on a probability of the set of OCR-output character attributes being erroneously interpreted by the OCR system;

See the rejection of Claim 1, second limitation.

determining a text passage error value for each OCR-output text passage based on a probability of the text passage being erroneously interpreted by the OCR system as determined using at least the OCR-output character error values; .

Fig. 5a-c

and comparing the determined text passage error value with an OCR-output text passage threshold error value to perform an OCR-output text passage triage decision.

See the rejection of Claim 1, third limitation.

13. The computer-implemented method of claim 12, wherein the probability of the set of OCR-output character attributes being erroneously interpreted by the OCR system is determined based on at least the selected set of OCR-output character attributes processed using the triage model.

See the rejection of Claim 2, second limitation.

14. The computer-implemented method of claim 12, wherein the plurality of OCR-output character attributes includes at least one of a character class, a confidence descriptor class, a language of the text passage, a text passage publication date, a typeface in which the text passage is printed, an image-based feature of an individual character image and metadata attached to the text passage.

See the rejection of Claim 5.

15. The computer-implemented method of claim 12, wherein the text passage to be triaged includes at least one of pages, characters, words, phrases, text-lines, sentences, paragraphs, columns of text, blocks of text, text articles, multi-page documents, collections of single-page documents and collections of multi-page documents.

See the rejection of Claim 6.

16. The computer-implemented method of claim 12, wherein the OCR-output text passage triage decision includes at least one of sending the OCR-output text passage directly to an end user without post-OCR processing, sending the OCR-output text passage through a post-OCR inspection and processing stage, and sending the original text passage image to be keyed in manually.

See the rejection of Claim 7.

17. The computer-implemented method of claim 16, wherein sending the OCR-output text passage through a post-OCR inspection and processing stage comprises:

determining at least one text passage error probability value for each OCR-output text passage as a correction operator detects and corrects an error in the OCR-output text passage;

See the rejection of Claim 10, first limitation.

and alerting the correction operator when the at least one text passage error probability value is improved so as to meet the OCR-output text passage threshold error value, wherein the text passage error probability value for each OCR-output text passage is based on a probability of the respective sets of OCR-output character attribute being erroneously interpreted by the system.

See the rejection of Claim 10, second limitation.

18. The computer-implemented method of claim 12, wherein determining a text passage error probability value for an OCR-output text passage comprises:

determining OCR-output text passage error probability values for a plurality of selected portions of the OCR-output text passage;

See rejection of claim 11, first limitation.

and arranging the plurality of selected portions of the OCR-output text passage based on the determined OCR-output text passage error probability values such that the selected portions having the highest OCR-output text passage error probability values are displayed first to the correction operator.

See rejection of claim 11, second limitation.

19. An OCR-output text passage triage system that triages a text passage outputted by an optical character recognition system, the OCR-output text passage including multiple OCR-output characters, each having at least one OCR-output character attribute, the system comprising:

an OCR-output text passage character accuracy determination circuit or routine that determines a character interpretation error value for individual OCR-output characters within the OCR-output text passage using a triage model;

See the rejection of Claim 1, second limitation..

an OCR-output text passage accuracy determination circuit or routine that determines at least one OCR-output text passage quality metric using the determined character interpretation error value and at least one statistical algorithm or model included in the triage model;

A quality metric as defined by the applicant is a text passage error value represented as a probability, that the entire OCR output text passage is erroneously interpreted by the OCR system, which has been shown to be another representation of the "confidence" disclosed by Bokser and is itself a statistical algorithm. See Claim 1, second limitation.

and an OCR-output text passage triage circuit or routine that performs one or more text passage triage decisions using the determined at least one OCR-output text passage quality metric and an OCR-output text passage threshold error rate value.

See above and see the rejection of Claim 1, third limitation..

20. The OCR-output text passage triage system of claim 19, wherein the triage model is a trained off-line triage model.

See the rejection of Claim 8.

21. The OCR-output text passage triage system of claim 19, wherein the OCR-output text passage threshold error rate value is included in a text passage error threshold operating point model.

Jamali discloses a model of triaging a passage of text which is synonymous with the applicant's "text passage error threshold operating point model", which is used to select a threshold operating point that will, with high confidence, satisfy customer-specified quality requirements while minimizing the labor needed to process document text passages that are not triaged. As can be seen in Jamali's disclosure the user-defined threshold is defined (col. 6, l. 8-11) and implemented (col. 6, l. 11-24) to satisfy customer-specified quality requirements with high confidence (col. 6, l. 25-48).

22. The OCR-output text passage triage system of claim 19, wherein the at least one OCR-output character attribute includes at least one of a character class, a confidence descriptor class, a language of the text passage, a text passage publication date, a typeface in which the text passage is printed, an image-based feature of an individual character image and metadata attached to the text passage.

See the rejection of Claim 5.

23. The OCR-output text passage triage system of claim 19, wherein the text passage to be triaged includes at least one of pages, characters, words, phrases, text-lines, sentences, paragraphs, columns of text, blocks of text, text articles, multi-page documents, collections of single-page documents and collections of multi-page documents.

See the rejection of Claim 6.

24. The OCR-output text passage triage system of claim 19, wherein the OCR-output text passage triage decision includes at least one of sending the OCR-output text passage directly to an end user without post-OCR rekeying or correction, sending the OCR-output text passage through a post-OCR inspection and correction stage, and sending the original text passage image to be completely keyed in manually.

See the rejection of Claim 7.

25. A computer-readable medium that provides instructions for triage of a text passage outputted by an optical character recognition system, the OCR-output text passage having multiple OCR-output characters, instructions, which when executed by a processor, cause the processor to perform operations comprising:

determining at least one OCR-output character attribute for each of the OCR-output characters in the OCR-output text passage;

See the rejection of Claim 1, first limitation.

determining an error rate for the OCR-output text passage using a triage model and the

determined OCR-output character attributes; and

See the rejection of Claim 1, second limitation.

comparing the determined error rate for the OCR-output text passage with an OCR-output

text passage threshold error rate to perform an OCR-output text passage triage decision.

See the rejection of Claim 1, third limitation.

26. The computer-readable medium of claim 25, wherein determining an error rate for the

OCR-output text passage comprises:

providing the OCR-output character attribute to the triage model;

See the rejection of Claim 2, first limitation.

determining a character interpretation error value for each OCR-output character based on

a probability of the at least one OCR-output character attribute being erroneously

interpreted by the system;

See the rejection of Claim 2, second limitation.

and determining a text passage error value based on the at least one character

interpretation error value determined for each OCR-output character.

See the rejection of Claim 2, third limitation.

27. The computer-readable medium of claim 26, further comprising:

determining a number representing a sum of OCR-output characters in the OCR-output text passage;

See the rejection of Claim 3, first limitation.

and dividing the text passage error value by the number representing the sum of OCR-output characters.

See the rejection of Claim 3, second limitation.

28. The computer-readable medium of claim 25, wherein determining at least one OCR-output character attribute for each OCR-output character comprises selecting the at least one OCR-output character attribute from a plurality of OCR-output character attributes.

See the rejection of Claim 4.

29. The computer-readable medium of claim 28, wherein the plurality of OCR-output character attributes includes at least one of a character class, a confidence descriptor class, a language of the text passage, a text passage publication date, a typeface in which the text passage is printed, an image-based feature of an individual character image and metadata attached to the text passage.

See the rejection of Claim 5.

30. The computer-readable medium of claim 25, wherein the text passage to be triaged includes at least one of pages, characters, words, phrases, text-lines, sentences, paragraphs, columns of text, blocks of text, text articles, multi-page documents, collections of single-page documents and collections of multi-page documents.

See the rejection of Claim 6.

31. The computer-readable medium of claim 25, wherein the OCR-output text passage triage decision includes at least one of sending the OCR-output text passage directly to an end user without post-OCR processing, sending the OCR-output text passage through a post-OCR inspection and processing stage, and sending the original text passage image to be keyed in manually.

See the rejection of Claim 7.

Conclusion

6.   The prior art made of record by not relied upon is Thompson et al. (U.S. Publication Number 2002/0103834) which discloses a system functionally similar to Jamali's in that it analyzes groups of character in a post-OCR process and determines accuracy information.

7.   **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Jonathan C. Schaffer whose telephone number is (571)272-0603. The examiner can normally be reached on 7:30am - 4:00pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Joseph Mancuso can be reached on (571)272-7695. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

JS
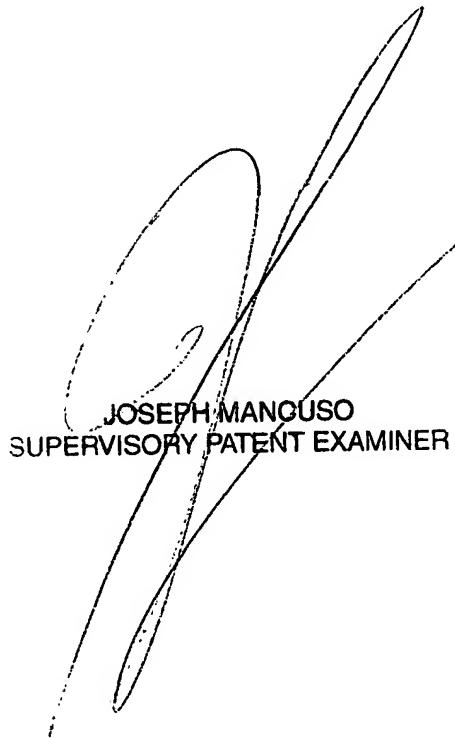
JOSEPH MANCUSO
SUPERVISORY PATENT EXAMINER